Global Journal of Engineering and Technology [GJET].
ISSN: **2583-3359** (Online)
Frequency: Monthly
Published By GSAR Publishers
Journal Homepage Link- https://gsarpublishers.com/journal-gjet-home/

OPEN ACCESS

Modeling and Simulation of Univariate and Multivariate analytics by applying Deep Learning and Machine Learning Application of Time Series application in the Neural Network Model.

BY

Ashraf Shahriar[1*], Md. Imrul Hassan[2], Md Atharshihab Biplob[3]

[1,2,3]Department of Electrical and Computer Engineering (ECE), ID:2025310050, 2025246050, 2025210050.
North South University, Plot # 15, Block # B, Bashundhara, Dhaka – 1229, Bangladesh.
Contact: e-mail: ashrafssuharto@gmail.com, imrul.hassan007@gmail.com, Biplob310@gmail.com

Article History

Received: 15/11/2024
Accepted: 27/11/2024
Published: 29/11/2024

Vol – 3 Issue – 11

PP: - 30-35

DOI:10.5281/zenodo.14249250

Abstract

The research addresses both univariate time series, where forecasts are made for a single variable over time, and multivariate time series, wherever multiple organized variables are used to develop predictive exactness. The data is split into training and testing sets by means of time series-specific practices such as sliding windows and expanding windows, confirming that chronological order is conserved and that models are authenticated effectually. Experimental consequences determine the applicability of deep learning models in accomplishing high accuracy for both short-term and long-term analytical tasks. The study compares the performance of univariate and multivariate tactics, highlighting the benefits of integrating multiple variables in refining forecast consistency (Adhikari and Ikeda, 2020). This research makes available a wide-ranging framework for applying machine learning and deep learning practices to time series prediction, offering insights into model selection, data preprocessing, and valuation approaches. The proposed approach validates substantial potential for real-world applications, enabling decision-makers to make informed forecasts and optimize procedures across various domains.

Keywords: *Univariate, Multivariate, Temperature, Humidity, Rainfall, Surface Soil Witness, Time Series, Chittagong, Chattogram, Bangladesh.*

I. INTRODUCTION

This study will make available an overview of how deep learning (DL) and machine learning (ML) can be functional to estimate the variable by means of univariate and multivariate time series analysis within the neural network context. By seeing multiple consistent aspects, models afford a comprehensive and thorough image of groundwater dynamics, empowering more precise forecasts and informed decision-making. Rainfall modeling and simulation are decisive tools for considerate and forecasting rainfall patterns. Particular temperature modeling supports in predicting weather, reviewing climate variation, and dealing agricultural performs (Lim and Zohren, 2021). More than a few machine learning algorithms can be castoff for research to find the above statements analysis, as well as Support Vector Regression (SVR), Random Forest (RF), K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU). The emphasis of the research has been fascinated on current enlargements, advancements, margins and insufficiencies of Advanced Neural Networking (ANN) tactics by using the Deep Learning system. The relevant and expected measurements' level data were employed to train and test the Neural Network. With the usage of the proficiency principles, mean square error (MSE), root mean square error (RMSE), and other metrics, each network structure's prediction accuracy was evaluated (R^2). Results have been demonstrated the time series forecasting in neural network (NN) model in the area of Chittagong (Chattogram) that is located in the southeastern zone of the country, at the mouth of the Karnaphuli river, in front of the Bay of Bengal in Bangladesh.

II. METHODOLOGY

Data Collection

Appropriate data was collected from NASA website for multivariate and BWDB for univariate time series analysis. The data would be collected at a right time interval (e.g., hourly, daily, monthly) based on the foretelling task. Multivariate models involve all variables to have corresponding timestamps. Testing models with artificial time series data generated based on known shapes. APIs make available real-time or historical data feeds for time series modeling. For univariate data entails a single variable

*Corresponding Author: Ashraf Shahriar

restrained over time. And for multivariate data involves collecting multiple connected variables over time. Scaling data (e.g., min-max scaling) supports neural networks converge quicker.

A. *Loss and accuracy function*

The choice of loss and accuracy functions is key for efficiently training and evaluating time series prediction models. The functions can be functional to both univariate and multivariate time series data, with slight amendments to handle numerous variables in the case of multivariate prediction (Fawaz, Forestier and Weber, 2019). The R-Squared is often apply to measure how well the estimates match the real values. It expresses how much of the modification in the time series data is enlightened by the model. A value of R^2=1 means perfect prediction, while R^2=0 suggests that the model doesn't explicate any of the variance. The R-squared (R^2) is a numerical formation. If the extent of loss function is high, it means algorithm is presentation a lot of alterations the consequence and desires be amended. Having a low accuracy calculation but a high loss would ensure that the model ensures big errors in the maximum of the dataset **(**Lundberg and Lee, 2017). But, if both loss and accuracy are low, it ensures the model makes small errors in the most of the dataset. Nevertheless, if they're both high, it shows big errors in some of the dataset.

Particulars	**Low Loss**	**High Loss**
Low accuracy	A lot of small errors	A lot of big errors
High accuracy	A few small errors	A few big errors

***Figure 1*: Low loss and high loss formation**

In time series prediction, if the model forecasts (Figure 1) values constantly close to the mean, the errors might be small, but the forecasts are not precise. A time series model trained with inadequate data might produce prophecies that are steadily off by large borders. A well-trained LSTM model in a univariate time series estimate where forecasts align closely with authentic values (Cheng, Li and Castelletti, 2020). In a multivariate time series, if the model captures the widely held of relationships but fails to predict extreme values precisely, the exactness may still be high while loss remains high.

A. *Split data for Training and Testing*

Splitting the data into training and testing sets is a crucial step when spread over deep learning and machine learning models to time series prediction. Unlike random splits often applied for other types of data, time series data involves careful reflection of its chronological structure to preserve the sequential order. The input data is split as training and testing (Figure:2): 65% for Training and 35% for Testing analysis:

65% Training Data	**35% Testing Data**

Figure 2: Split data for training and testing

A. *Rolling windows with adjusting training size and Constant training size*

When applying rolling windows to time series modeling with deep learning or machine learning, we can use adjusting training size or constant training size methods to iteratively train and test the typical. These tactics are considered to appraise the performance of a model over time, ensuring adaptableness to the active patterns and trends in the dataset. Rolling window procedures split the time series data into training and testing sets by sliding the window frontward step-by-step through the dataset. Simulating real-world projecting tasks where the forthcoming must be predicted based on historical dataset (Karthikeyan, Khosa and Singh, 2020). By integrating rolling windows with either altering training size or constant training size, we can figure robust and flexible models for univariate and multivariate time series data. The choice of tactic depends on the features of dataset and the precise goals of the modeling task.

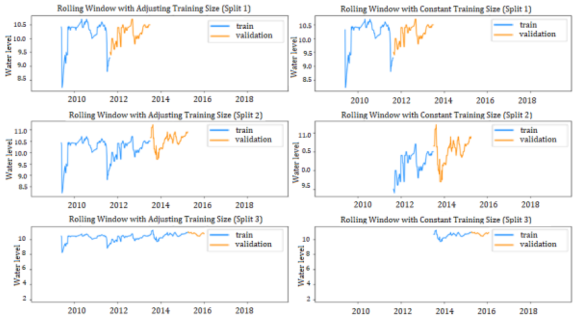


***Figure 3:* Rolling windows with adjusting training size and Constant training size**

Observations: The training set (Figure 3) raises over time as the window inflates to include more historical data. Each split adds more historic data to the training set, while the validation set remains static in size. The blue line signifies the rising training dataset, and the orange line signifies the fixed authentication dataset. The training set size remains constant, directing on a fixed antique window of data. The training window changes forward in time, while the validation set also moves consequently. Both training and validation sizes are kept reliable across splits.

*Corresponding Author: Ashraf Shahriar

Flowchart 1: **Process flow and Data organization**

D. *Time series prediction to forecast the future groundwater levels, considering factors such as water table depth, parapet height, and geographical directs comparing the latitude and Longitude perspective.*

Sl	District	Well Id	Water Level	If R l Parapet (M)	Parapet Height (M)	Depth (M)	Latitude	Longitude
1	Chattogram	GT1541007	10.42	4.6	0.68	4.31	22.32771	91.80986
2	Chattogram	GT1541007	10.32	4.6	0.68	4.31	22.32771	91.80986
3	Chattogram	GT1541007	8.22	4.6	0.68	4.31	22.32771	91.80986
4	Chattogram	GT1541007	8.32	4.6	0.68	4.31	22.32771	91.80986
5	Chattogram	GT1541007	9.30	4.6	0.68	4.31	22.32771	91.80986

***Table 1*: For depth, parapet height, and geographical directs**

Application and Summary of SVR, RF, KNN, LSTM, GRU, LSTM+GRU algorithms. This graph provides a visual comparison of the performance of different machine learning algorithms in predicting GWL in Chittagong applying the Train and Test perspective.

Algorithms	Train RMSE	Test RMSE	Train MSE	Test MSE	Train MAE	Test MAE	Train VRS	Test VRS	Train R2 Score	Test R2 Score	Train MGD	Test MGD	Train MPD	Test MPD
SVR	0.438896	1.808098	0.192630	3.269220	0.370113	1.011549	0.042243	0.257900	-0.320398	0.257857	0.001891	0.073065	0.019078	0.467049
RF	0.126041	1.918553	0.015886	3.680844	0.091845	0.743203	0.891279	0.192565	0.891106	0.164414	0.000158	0.079264	0.001583	0.517886
KNN	0.335650	1.985201	0.112661	3.941022	0.251517	0.829393	0.236772	0.123661	0.227754	0.105351	0.001107	0.082007	0.011163	0.544601
LSTM	0.395867	1.968437	0.156711	3.874743	0.325369	1.025143	0.091973	0.120480	-0.074186	0.120397	0.001531	0.080712	0.015481	0.534984
GRU	0.363580	1.790348	0.132190	3.205344	0.254162	0.719432	0.207420	0.284985	0.093891	0.272357	0.001293	0.073485	0.013065	0.465498
LSTM+GRU	0.430805	2.007958	0.185593	4.031895	0.302637	0.751406	0.097552	0.122972	-0.272161	0.084722	0.001788	0.082523	0.018208	0.551782

Table 2: performance of different machine learning algorithms

III. MODELING AND SIMULATION

A. Univariate Time Series Forecasting for Groundwater Level (GWL):

Ensure data consistency by handling missing values and smoothing out noise. Apply scaling (e.g., Min-Max Scaling or Standardization) for better convergence during model training. Use stationarity tests to check if the time series is stationary. Apply differencing or transformations if needed. Evaluate performance using metrics.

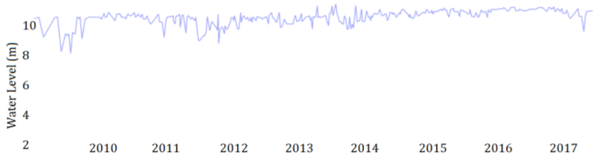


Figure 4: **GWL chart, water level**

Observations: The variations seem (Figure 4) to developed less extreme in the later years (2014-2017), with the line becoming slightly more stable and reliably higher. Sharp drop near the end of 2017, but it rapidly improves.

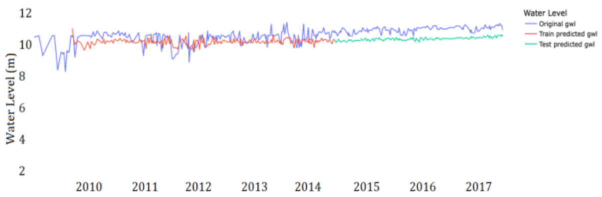


Figure 5: **Original Vs predicted GWL (SVR)**

Observations: There's a visible (Figure 5) enhancement in the model's forecasts after 2014, where the projected lines more closely follow the original data. Water level usually stays between 8 and 12 meters all over the period.

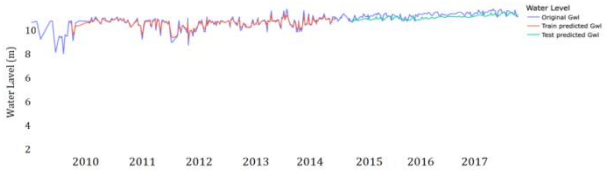


Figure 6: Original Vs predicted GWL (RF)

Observations: The model's forecasts seem (Figure 6) to progress after 2014, with closer alignment to the original data. The water level mostly stays between 8 and 11 meters all over the period. There's a visible dip in water levels around 2010, captured by all three lines.

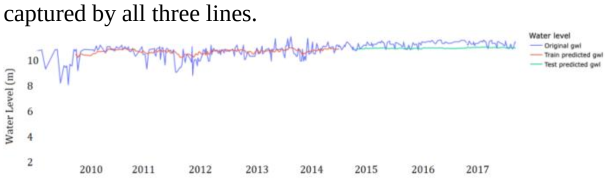
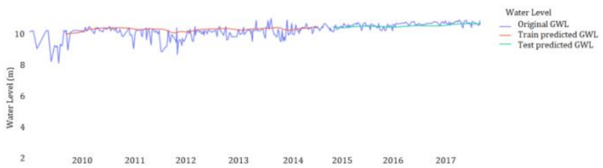


Figure 7: Original Vs predicted GWL (LSTM)

Observations: The model seems (Figure 7) to capture the overall trend well but misses some of the extreme highs and lows in the original data. Towards the later years (2015-2017), the predicted lines align more closely with the original data, suggesting possibly improved model performance or more stable water levels.



*Corresponding Author: Ashraf Shahriar

Figure 8: **Original Vs predicted GWL (LSTM+GRU)**

Observations: The blue line (original data) varies but generally shows (Figure 8) stability over the years, with minor cyclical or irregular variations. The red line overlaps with the blue line in the training phase, signifying the model has captured the trends in the training data effectively. The green line in the test phase also strictly follows the blue line, signifying that the model simplifies well to invisible data. Overall, the forecast GWL values (both for training and testing) align thoroughly with the original GWL, reflecting a good fit of the analytical model to the data.

B. Multivariate Time Series Forecasting for Groundwater Level, Rainfall, Temperature, Root and Surface Soil Witness, Depth to Groundwater level.

Multivariate time series involves analyzing multiple variables simultaneously, considering their interactions and joint impact on the target variable. In addition to neural networks, traditional machine learning models can simulate and forecast time series, especially when the problem is less complex or data is limited (Abdollahi, Bazrafshan and Razmjooy, 2020). Assess interdependence by calculating cross-correlation between predicted and actual variables.

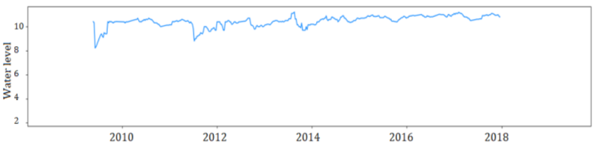


Figure 9: Water Level analysis

Observations: There's a notable sharp (Figure 9) drop near the beginning of 2010, but it quickly recovers. After 2010, the water level remains mostly consistent with minor variations.

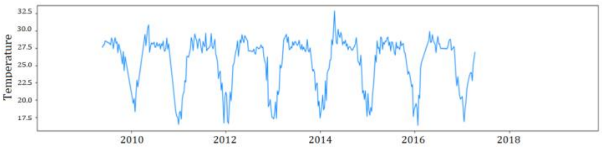


Figure 10: Temperature analysis

Observations: Highest peaks (Figure 10) reach around 30-32 degrees, while lowest points dip to 17-20 degrees. This cyclical pattern repeats consistently each year throughout the time period shown.

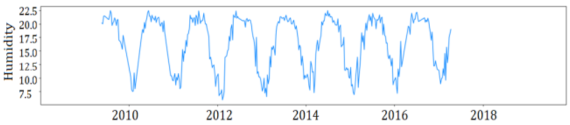


Figure 11: Humidity analysis

Observations: The humidity (Figure 11) levels generally fluctuate between 10 and 22.5 units. Consistent of annual patterns, representing seasonal changes in humidity.

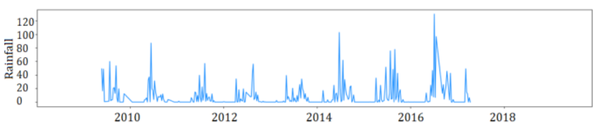


Figure 12: Rainfall analysis

Observations: Displays (Figure 12) sporadic spikes of varying heights throughout the time period. Most of the time, rainfall is low or zero, with occasional sharp increases. The highest spike reaches about 120 units, occurring around 2016.

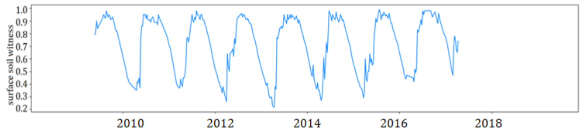


Figure 13: Surface root witness analysis

Observations: The cycles appear (Figure 13) be seasonal, with peaks occurring roughly. Variation in the patterns between years, but overall cyclical nature remains consistent.

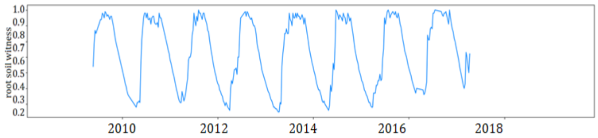


Figure 14: Root soil witness

Observations: The graph exhibits (Figure 14) a strong periodicity, with peaks occurring approximately once every year, suggesting an annual cycle. The peaks and troughs are consistent in amplitude and timing, which might indicate a naturally occurring phenomenon like annual rainfall, temperature fluctuations, or crop cycles. Near the end of 2017 and the start of 2018, the pattern slightly deviates, with a dip or regularity that could signify a disruption in the typical cycle.

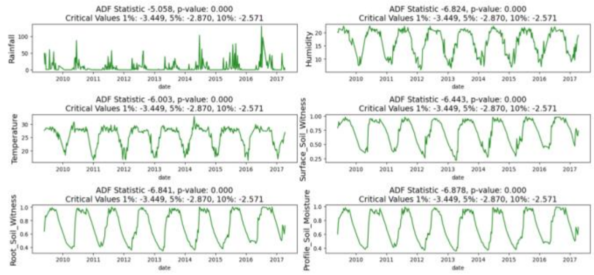


Figure 15: **Numerical time series graphs**

Observations: This image shows 9Figure 15) a set of six time series graphs, each representing different environmental variables from 2010 to 2017. Each graph includes an Augmented Dickey-Fuller (ADF) statistic and p-value, along with critical values, indicating tests for stationarity. All graphs have negative ADF statistics and p-values of 0.000, suggesting that these time series are stationary (do not have unit roots). The critical values provided are the same for all graphs, indicating consistent testing parameters. This compilation of graphs provides a comprehensive view of various interrelated environmental factors over time, useful for studying climate patterns, agricultural conditions, or hydrological cycles.

IV. FINDINGS AND RECOMMENDATIONS

The DL and ML models could categorize the lagged relationships and seasonality patterns that should be

*Corresponding Author: Ashraf Shahriar

influenced the behavior of the variables over time. The study could expose nonlinear relationships and interactions between the ecological variables such as neural networks, are well-suited for capturing complex, nonlinear changing aspects in multivariate time series data (Mojid, Parvez, Mainuddin and Hodgson, 2019). The study area might validate the superior calculation accuracy of DL and ML models associated to the traditional statistical approaches for time series forecasting. Uncertainty breakdown could afford to the decision-makers with a more nuanced considerate of the reliability and restrictions of predicting consequences. The interdisciplinary nature of the study could foster relationships among hydrologists, climatologists, agronomists, and data scientists to address the complex water-related challenges from multiple viewpoints (Aishwarya and Vasudevan, 2023). The models confirmed varying degrees of correctness in forecasting the groundwater levels, with some outperforming others in the certain circumstances. Implementing multivariate time series estimating models that concurrently forecast multiple variables of interest that allows for capturing the complex interactions and response loops amid ecological variable quantity. Normalizing the input variables to ensure that they are on a similar scale and facilitate convergence during training. Reflects using collective approaches to combine estimates from multiple models or model alternatives. Measuring the uncertainty associated with the model forecasts and evaluate the sensitivity of outcomes to differences in input parameters and model assumptions. Authenticate the predicting models using independent datasets and real-world observations to govern generalizability and applicability in effective backgrounds.

V. RESULTS

A. *Loss Score*

Loss Score shows, how widely loss originate where are three loss functions in the six of our algorithms. Most losses originate to the RMSE amongst the algorithms. The number oscillated amongst 1.22 and 3.80 for RMSE. By distinction, the MAE and the MSE acknowledged scarcer losses correspondingly. The quantity of wounded which originate the MSE vacillated from 2.91 to 3.84 from SVR to RF. The fatalities loss diminished in LSTM and then augmented in LSTM+GRU. So, the tendency for the RMSE was the almost similar. There was a fluctuation rise in losses from SVR to KNN. The loss decreased in LSTM and then increased in LSTM+GRU.

B. *Accuracy Score*

Accuracy Scores stretches of evidence approximately how much accurateness found two accuracy purposes in the six of the total algorithms. Test R2 Score is the actual accuracy score of the algorithms. As can be seen from the graph, there were different trends for Train R2 Score and Test R2 Score. The quantity of Train R2 Score increased -0.220 to 0.90 for SVR and RF harmoniously. Subsequently Train R2 Slash progressively deteriorated to 0.226 in the LSTM+GRU algorithms. On the other hand, the Test R2 Score fluctuated between 0.223 to 0.226 for SVR and GRU+LSTM respectively. The highest accuracies are found in RF and LSTM algorithms which is about 0.226 for both.

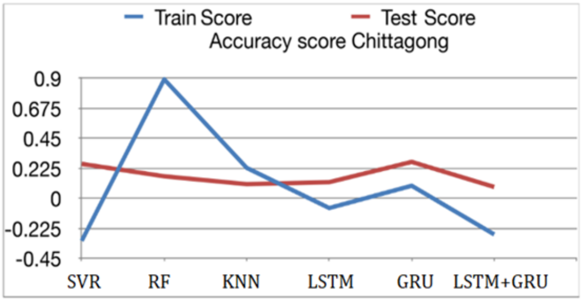


Figure 16: Accuracy score

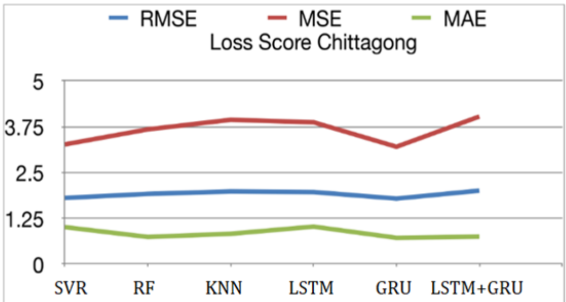


Figure 17: Loss Score

C. *Accuracy Score of heatmap:*

To illustrate the accuracy score of a heat map in a framework as, several issues need to be measured, as heat maps are visualization tools rather than direct assessing metrics.

(I) ***Workflow to Determine Accuracy:***

- Train the Neural Network Model
- Illustrate the Model Performance
- Quantify Accuracy of Predictions

(II) ***Specific Terms and applications:***

- To ensure the feature Importance Metrics
- Correlation Scores among the application of Algorithms
- Interpretability should align with expected domain knowledge

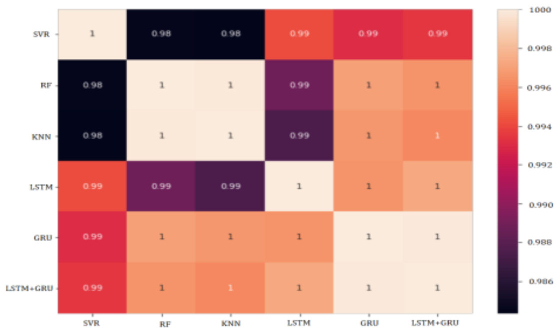


***Figure 18*: Accuracy Score Heatmap correlation of algorithms**

Observations: This heatmap suggests (Figure 18) that models are producing very similar predictions, with only slight variations. High correlations indicate the different algorithms are capturing similar patterns, which could be interpreted as a sign of robust modeling or, alternatively, might suggest that

*Corresponding Author: Ashraf Shahriar

simpler models could be sufficient if all perform so similarly. Most of the correlations are very high, ranging from 0.98 to 1.0, indicating strong positive relationships between the different models' predictions (Fawaz and Weber, 2019).

VI. CONCLUSION

By combining deep learning and machine learning analytics with univariate and multivariate time series forecasting techniques, this can advance the analytical models that capture the complex interactions between groundwater dynamics, rainfall, temperature, soil moisture, and other ecological variable quantity (Gharbi and Bouaziz, 2023). These models can help stakeholders anticipate fluctuations in water accessibility, improve reserve allocation, and moderate the influences of climate unpredictability and variation. Train the particular models by means of the training data and tune hyper restrictions using the authentication set. Applications of univariate analytics Predicting stock prices, energy consumption, or sales trends and for the multivariate are Weather forecasting, demand planning in supply chains, or predicting disease progression in healthcare. This article to serve as an insightful and comprehensive resource for researchers and experts in the area.

REFERENCES

1. Abdollahi, A., Bazrafshan, J., & Razmjooy, N. (2020). *A deep learning method for temperature forecasting based on Bi-LSTM.* Applied Soft Computing, 97, 106782.
2. Adhikari, U., & Ikeda, M. (2020). *XGBoost and LSTM model for multivariate time series forecasting of groundwater level. Water*, 12(11), 3082. DOI.
3. Aishwarya, R., & Vasudevan, V. (2023). *A comprehensive review of multi-source data integration for groundwater modeling. Journal of Hydrology*, 617, 128812. DOI.
4. Cheng, L., Li, G., & Castelletti, A. (2020). GWL forecasting with LSTM networks: *A study in agricultural water management. Hydrological Sciences Journal*, 65(9), 1505-1516.
5. Fawaz, H. I., Forestier, G., Weber, J., et al. (2019). Deep learning for time series classification: *A review. Data Mining and Knowledge Discovery*, 33(4), 917-963.
6. Gharbi, S., & Bouaziz, M. (2023). *Real-time data assimilation for GWL prediction using machine learning.* Water Resources Management, 37(5), 1461-1478. DOI.
7. Karthikeyan, L., Khosa, R., & Singh, V. P. (2020). *Deep learning models for soil moisture retrieval from remote sensing data*: A review. Environmental Earth Sciences, 79(7), 193.
8. Lim, B., & Zohren, S. (2021). Time series forecasting with deep learning: A survey. Philosophical Transactions Royal Society: Mathematical, Physical and Engineering Sciences, 379(2194), 20200209.
9. Lundberg, S. M., & Lee, S. I. (2017). *A unified approach to interpreting model predictions.* In Advances in Neural Information Processing Systems (pp. 4765-4774).
10. Mojid, M.A., Parvez, M.F., Mainuddin, M. and Hodgson, G., (2019), Water Table Trend- *A Sustainability of GWL Development in North-West Bangladesh*, Water, Vol 11, pp-1182.

*Corresponding Author: Ashraf Shahriar